

(1) APPLICANT: SHIONOGI SEIYAKU KABUSHIKI KAISHA

(iii) NUMBER OF SEQUENCES: 2

(A) ADDRESSEE: SHIONOGI SEIYAKU KABUSHIKI  
KAISHA

(B) STREET: 3-1-8, Dosho-machi

(C) CITY: Chuo-ku

(D) STATE: Osaka

(E) COUNTRY: Japan

(F) ZIP: 541

(A) MEDIUM TYPE: Diskette. 3.50 inch.  
1024 kb storage

(B) COMPUTER: EPSON

(C) OPERATING SYSTEM: MS-DOS ver. 2.11

(D) SOFTWARE: Wordstar ver. 5.0

(A) APPLICATION NUMBER: JP 3-172828

(B) FILING DATE: 12-July-1991

[illegible]

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 4105 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: human

(vii) DIRECT ORIGINAL SOURCE:

- (B) CLONE: phETIR

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 485 .. 1768
- (C) IDENTIFICATION METHOD: by experiment
- (A) NAME/KEY: signal peptide
- (B) LOCATION: 485 .. 544
- (C) IDENTIFICATION METHOD: similarity to other signal sequences
- (A) NAME/KEY: mat peptide
- (B) LOCATION: 545 .. 1768
- (C) IDENTIFICATION METHOD: similarity to other signal sequences

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

GAATTCGCGG CCGCCTCTTG CGGTCCCAGA GTGGAGTGA AGGTCTGGAG CTTTGGGAGG 60

AGACGGGGAG GACAGACTGG AGGCGTGTTT CTCCGGAGTT TTCTTTTTCG TGCGAGCCCT 120

Sub B1

CGCGCGCGCG TACAGTCATC CCGCTGGTCT GACGATTGTG GAGAGGCGGT GGAGAGGCTT 180

CATCCATCCC ACCCGGTCGT CGCCGGGGAT TGGGGTCCCA GCGACACCTC CCCGGGAGAA 240

GCAGTGCCCA GGAAGTTTTC TGAAGCCGGG GAAGCTGTGC AGCCGAAGCC GCCGCCGCGC 300

CGGAGCCCGG GACACCGGCC ACCCTCGCGG CCACCCACCC TCGCTTTCTC CGGCTTCCTC 360

TGGCCCAGGC GCCGCGCGGA CCCGGCAGCT GTCTGCGCAC GCCGAGCTCC ACGGTGAAAA 420

AAAAAGTGAA GGTGTAAAAG CAGCACAAGT GCAATAAGAG ATATTTCTCT AAATTTGCCT 480

CAAG ATG GAA ACC CTT TGC CTC AGG GCA TCC TTT TGG CTG GCA CTG GTT 529

Met Glu Thr Leu Cys Leu Arg Ala Ser Phe Trp Leu Ala Leu Val  
 -20 -15 -10

GGA TGT GTA ATC AGT GAT AAT CCT GAG AGA TAC AGC ACA AAT CTA AGC 577

Gly Cys Val Ile Ser Asp Asn Pro Glu Arg Tyr Ser Thr Asn Leu Ser  
 -5 -1 1 5 10

AAT CAT GTG GAT GAT TTC ACC ACT TTT CGT GGC ACA GAG CTC AGC TTC 625

Asn His Val Asp Asp Phe Thr Thr Phe Arg Gly Thr Glu Leu Ser Phe  
 15 20 25

CTG GTT ACC ACT CAT CAA CCC ACT AAT TTG GTC CTA CCC AGC AAT GGC 673

Leu Val Thr Thr His Gln Pro Thr Asn Leu Val Leu Pro Ser Asn Gly  
 30 35 40

TCA ATG CAC AAC TAT TGC CCA CAG CAG ACT AAA ATT ACT TCA GCT TTC 721,

Subbi  
Ser Met His Asn Tyr Cys Pro Gln Gln Thr Lys Ile Thr Ser Ala Phe  
45 50 55

AAA TAC ATT AAC ACT GTG ATA TCT TGT ACT ATT TTC ATC GTG GGA ATG 769

Lys Tyr Ile Asn Thr Val Ile Ser Cys Thr Ile Phe Ile Val Gly Met  
60 65 70 75

CTG GGG AAT GCA ACT CTG CTC AGG ATC ATT TAC CAG AAC AAA TGT ATG 817

Val Gly Asn Ala Thr Leu Leu Arg Ile Ile Tyr Gln Asn Lys Cys Met  
80 85 90

AGG AAT GGC CCC AAC GCG CTG ATA GCC AGT CTT GCC CTT GGA GAC CTT 865

Arg Asn Gly Pro Asn Ala Leu Ile Ala Ser Leu Ala Leu Gly Asp Leu  
95 100 105

ATC TAT GTG GTC ATT GAT CTC CCT ATC AAT GTA TTT AAG CTG CTG GCT 913

Ile Tyr Val Val Ile Asp Leu Pro Ile Asn Val Phe Lys Leu Leu Ala  
110 115 120

GGG CGC TGG CCT TTT GAT CAC AAT GAC TTT GGC GTA TTT CTT TGC AAG 961

Gly Arg Trp Pro Phe Asp His Asn Asp Phe Gly Val Phe Leu Cys Lys  
125 130 135

CTG TTC CCC TTT TTG CAG AAG TCC TCG GTG GGG ATC ACC GTC GTC AAC 1009

Leu Phe Pro Phe Leu Gln Lys Ser Ser Val Gly Ile Thr Val Leu Asn  
140 145 150 155

CTC TGC GCT CTT AGT GTT GAC AGG TAC AGA GCA GTT GCC TCC TGG AGT 1057

Sub B1

Leu	Cys	Ala	Leu	Ser	Val	Asp	Arg	Tyr	Arg	Ala	Val	Ala	Ser	Trp	Ser		
				160					165					170			
CGT	GTT	CAG	GGA	ATT	GGG	ATT	CCT	TTG	GTA	ACT	GCC	ATT	GAA	ATT	GTC	1105	
Arg	Val	Gln	Gly	Ile	Gly	Ile	Pro	Leu	Val	Thr	Ala	Ile	Glu	Ile	Val		
			175					180					185				
TCC	ATC	TGG	ATC	CTG	TCC	TTT	ATC	CTG	GCC	ATT	CCT	GAA	GCG	ATT	GGC	1153	
Ser	Ile	Trp	Ile	Leu	Ser	Phe	Ile	Leu	Ala	Ile	Pro	Glu	Ala	Ile	Gly		
			190					195				200					
TTC	GTC	ATG	GTA	CCC	TTT	GAA	TAT	AGG	GGT	GAA	CAG	CAT	AAA	ACC	TGT	1201	
Phe	Val	Met	Val	Pro	Phe	Glu	Tyr	Arg	Gly	Glu	Gln	His	Lys	Thr	Cys		
			205					210				215					
ATG	CTC	AAT	GCC	ACA	TCA	AAA	TTC	ATG	GAG	TTC	TAC	CAA	GAT	GTA	AAG	1249	
Met	Leu	Asn	Ala	Thr	Ser	Lys	Phe	Met	Glu	Phe	Tyr	Gln	Asp	Val	Lys		
220					225				230						235		
GAC	TGG	TGG	CTC	TTC	GGG	TTC	TAT	TTC	TGT	ATG	CCC	TTG	GTG	TGC	ACT	1297	
Asp	Trp	Trp	Leu	Phe	Gly	Phe	Tyr	Phe	Cys	Met	Pro	Leu	Val	Cys	Thr		
			240						245					250			
GCG	ATC	TTC	TAC	ACC	CTC	ATG	ACT	TGT	GAG	ATG	TTG	AAC	AGA	AGG	AAT	1345	
Ala	Ile	Phe	Tyr	Thr	Leu	Met	Thr	Cys	Glu	Met	Leu	Asn	Arg	Arg	Asn		
			255					260					265				

Sub B1

GGC AGC TTG AGA ATT GCC CTC AGT GAA CAT CTT AAG CAG CGT CGA GAA	1393
Gly Ser Leu Arg Ile Ala Leu Ser Glu His Leu Lys Gln Arg Arg Glu	
270 275 280	
GTG GCA AAA ACA GTT TTC TGC TTG GTT GTA ATT TTT GCT CTT TGC TGG	1441
Val Ala Lys Thr Val Phe Cys Leu Val Val Ile Phe Ala Leu Cys Trp	
285 290 295	
TTC CCT CTT CAC TTA AGC CGT ATA TTG AAG AAA ACT GTG TAT AAC GAA	1489
Phe Pro Leu His Leu Ser Arg Ile Leu Lys Lys Thr Val Tyr Asn Glu	
300 305 310 315	
ATG GAC AAG AAC CGA TGT GAA TTA CTT AGT TTC TTA CTG CTC ATG GAT	1537
Met Asp Lys Asn Arg Cys Glu Leu Leu Ser Phe Leu Leu Leu Met Asp	
320 325 330	
TAC ATC GGT ATT AAC TTG GCA ACC ATG AAT TCA TGT ATA AAC CCC ATA	1585
Tyr Ile Gly Ile Asn Leu Ala Thr Met Asn Ser Cys Ile Asn Pro Ile	
335 340 345	
GCT CTG TAT TTT GTG AGC AAG AAA TTT AAA AAT TGT TTC CAG TCA TGC	1633
Ala Leu Tyr Phe Val Ser Lys Lys Phe Lys Asn Cys Phe Gln Ser Cys	
350 355 360	
CTC TGC TGC TGC TGT TAC CAG TCC AAA AGT CTG ATG ACC TCG GTC CCC	1681
Leu Cys Cys Cys Cys Tyr Gln Ser Lys Ser Leu Met Thr Ser Val Pro	
365 370 375	

Sub B<sup>1</sup>

ATG AAC GGA ACA AGC ATC CAG TGG AAG AAC CAC GAT CAA AAC AAC CAC 1729

Met Asn Gly Thr Ser Ile Gln Trp Lys Asn His Asp Gln Asn Asn His  
380 385 390 395

AAC ACA GAC CGG AGC AGC CAT AAG GAC AGC ATG AAC TGACCACCCT 1775

Asn Thr Asp Arg Ser Ser His Lys Asp Ser Met Asn  
400 405

TAGAAGCACT CCTCGGTACT CCCATAATCC TCTCGGAGAA AAAAATCACA AGGCAACTGT 1835

GACTCCGGGA ATCTCTTCTC TGATCCTTCT TCCTTAATTC ACTCCCACAC CCAAGAAGAA 1895

ATGCTTTCCA AAACCGCAAG GTAGACTGGT TTATCCACCC ACAACATCTA CGAATCGTAC 1955

TTCTTTAATT GATCTAATTT ACATATTCTG CGTGTTGTAT TCAGCACTAA AAAATGGTGG 2015

GAGCTGGGGG AGAATGAAGA CTGTTAAATG AAACCAGAAG GATATTTACT ACTTTTGCAT 2075

GAAAATAGAG CTTTCAAGTA CATGGCTAGC TTTTATGGCA GTTCTGGTGA ATGTTCAATG 2135

GGAAGTGGTC ACCATGAAAC TTTAGAGATT AACGACAAGA TTTTCTACTT TTTTAAAGTG 2195

ATTTTTTGTC CTTAGCCAA ACACAATATG GGCTCAGGTC ACTTTTATTT GAAATGTCAT 2255

TTGGTGCCAG TATTTTTTAA CTGCATAATA GCCTAACATG ATTATTTGAA CTTATTTACA 2315

CATAGTTTGA AAAAAAAG ACAAAAATAG TATTCAGGTG AGCAATTAGA TTAGTATTTT 2375

CCACGTCCT ACTTATTTTT TTAAACACA AATTCTAAAG CTACAACAA TACTACAGGC 2435

CCTTAAAGCA CAGTCTGATG ACACATTGG CAGTTTAATA GATGTTACTC AAAGAATTTT 2495

TTAAGAACTG TATTTTATTT TTAAATGGT GTTTTATTAC AAGGGACCTT GAACATGTTT 2555

SubB1

TGTATGTTAA ATTCAAAAGT AATGCTTCAA TCAGATAGTT CTTTTTCACA AGTTCAATAC 2615

TGTTTTTCAT GTAAATTTTG TATGAAAAAT CAATGTCAAG TACCAAAATG TTAATGTATG 2675

TGTCATTTAA CTCTGCCTGA GACTTTCAGT GCACTGTATA TAGAAGTCTA AAACACACCT 2735

AAGAGAAAAA GATCGAATTT TTCAGATGAT TCGGAAATTT TCATTCAGGT ATTTGTAATA 2795

GTGACATATA TATGTATATA CATATCACCT CCTATTCTCT TAATTTTTGT TAAAATGTTA 2855

ACTGGCAGTA AGTCTTTTTT GATCATTTCC TTTTCCATAT AGGAAACATA ATTTTGAAGT 2915

GGCCAGATGA GTTTATCATG TCAGTGAAAA ATAATTACCC ACAAATGCCA CCAGTAACTT 2975

AACGATTCTT CACTTCTTGG GGTTTTCAGT ATGAACCTAA CTCCCCACCC CAACATCTCC 3035

CTCCCACATT GTCACCATTT CAAAGGGCCC ACAGTGACTT TTGCTGGGCA TTTTCCCAGA 3095

TGTTTACAGA CTGTGAGTAC AGCAGAAAAT CTTTTACTAG TGTGTGTGTG TATATATATA 3155

AACAATTGTA AATTTCTTTT AGCCCATTTT TCTAGACTGT CTCTGTGGAA TATATTTGTG 3215

TGTGTGATAT ATGCATGTGT GTGATGGTAT GTATGGATTT AATCTAATCT AATAATTGTG 3275

CCCCGCAGTT GTGCCAAAGT GCATAGTCTG AGCTAAAATC TAGGTGATTG TTCATCATGA 3335

CAACCTGCCT CAGTCCATTT TAACCTGTAG CAACCTTCTG CATTCAATAA TCTTGTAATC 3395

ATGTTACCAT TACAAATGGG ATATAAGAGG CAGCGTGAAA GCAGATGAGC TGTGGACTAG 3455

CAATATAGGG TTTTGTTTGG TTGGTTGGTT TGATAAAGCA GTATTTGGGG TCATATTGTT 3515

TCCTGTGCTG GAGCAAAAGT CATTACACTT TGAAGTATTA TATTGTTCTT ATCCTCAATT 3575

CAATGTGGTG ATGAAATTGC CAGGTTGTCT GATATTTCTT TCAGACTTCG CCAGACAGAT 3635



SubB1

TGCTGATAAT AAATTAGGTA AGATAATTTG TTGGGCCATA TTTTAGGACA GGTAATAATA 3695

CATCAGGTTC CAGTTGCTTG AATTGCAAGG CTAAGAAGTA CTGCCCTTTT GTGTGTTAGC 3755

AGTCAAATCT ATTATTCCAC TGGCGCATCA TATGCAGTGA TATATGCCTA TAAATAAAGC 3815

CATAGGTTC CACCATTTTG TTTAGACAAT TGTCTTTTTT TCAAGATGCT TTGTTTCTTT 3875

CATATGAAAA AAATGCATTT TATAAATTCA GAAAGTCATA GATTTCTGAA GCGTCAACG 3935

TGCATTTTAT TTATGGACTG GTAAGTAACT GTGGTTTACT AGCAGGAATA TTTCCAATTT 3995

CTACCTTTAC TACATCTTTT CAACAAGTAA CTTTGTAGAA ATGAGCCAGA AGCCAAGGCC 4055

CTGAGTTGGC AGTGGCCCAT AAGTGTAATA TAAAAGTTTA CAGAAACCTT 4105

(2) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 4301 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: human

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 238 .. 1566
- (C) IDENTIFICATION METHOD: by experiment

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

GAGACATTCC GGTGGGGGAC TCTGGCCAGC CCGAGCAACG TGGATCCTGA GAGCACTCCC 60  
AGGTAGGCAT TTGCCCCGGT GGGACGCCTT GCCAGAGCAG TGTGTGGCAG GCCCCCGTGG 120  
AGGATCAACA CAGTGGCTGA ACACTGGGAA GGAAGTGGTA CTTGGAGTCT GGACATCTGA 180  
AACTTGGCTC TGAAACTGCG GAGCGGCCAC CGGACGCCTT CTGGAGCAGG TAGCAGC 237  
ATG CAG CCG CCT CCA AGT CTG TGC GGA CGC GCC CTG GTT GCG CTG GTT 285  
Met Gln Pro Pro Pro Ser Leu Cys Gly Arg Ala Leu Val Ala Leu Val  
1 5 10 15  
CTT GCC TGC GGC CTG TCG CGG ATC TGG GGA GAG GAG AGA GGC TTC CCG 333  
Leu Ala Cys Gly Leu Ser Arg Ile Trp Gly Glu Glu Arg Gly Phe Pro  
20 25 30  
CCT GAC AGG GCC ACT CCG CTT TTG CAA ACC GCA GAG ATA ATG ACG CCA 381  
Pro Asp Arg Ala Thr Pro Leu Leu Gln Thr Ala Glu Ile Met Thr Pro  
35 40 45  
CCC ACT AAG ACC TTA TGG CCC AAG GGT TCC AAC GCC AGT CTG GCG CGG 429  
Pro Thr Lys Thr Leu Trp Pro Lys Gly Ser Asn Ala Ser Leu Ala Arg  
50 55 60  
TCG TTG GCA CCT GCG GAG GTG CCT AAA GGA GAC AGG ACG GCA GGA TCT 477  
Ser Leu Ala Pro Ala Glu Val Pro Lys Gly Asp Arg Thr Ala Gly Ser  
65 70 75 80  
CCG CCA CGC ACC ATC TCC CCT CCC CCG TGC CAA GGA CCC ATC GAG ATC 525

Subbi

Pro	Pro	Arg	Thr	Ile	Ser	Pro	Pro	Pro	Cys	Gln	Gly	Pro	Ile	Glu	Ile		
				85					90					95			
AAG	GAG	ACT	TTC	AAA	TAC	ATC	AAC	ACG	GTT	GTG	TCC	TGC	CTT	GTG	TTC		573
Lys	Glu	Thr	Phe	Lys	Tyr	Ile	Asn	Thr	Val	Val	Ser	Cys	Leu	Val	Phe		
			100					105					110				
GTG	CTG	GGG	ATC	ATC	GGG	AAC	TCC	ACA	CTT	CTG	AGA	ATT	ATC	TAC	AAG		621
Val	Leu	Gly	Ile	Ile	Gly	Asn	Ser	Thr	Leu	Leu	Arg	Ile	Ile	Tyr	Lys		
			115				120					125					
AAC	AAG	TGC	ATG	CGA	AAC	GGT	CCC	AAT	ATC	TTG	ATC	GCC	AGC	TTG	GCT		669
Asn	Lys	Cys	Met	Arg	Asn	Gly	Pro	Asn	Ile	Leu	Ile	Ala	Ser	Leu	Ala		
			130			135						140					
CTG	GGA	GAC	CTG	CTG	CAC	ATC	GTC	ATT	GAC	ATC	CCT	ATC	AAT	GTC	TAC		717
Leu	Gly	Asp	Leu	Leu	His	Ile	Val	Ile	Asp	Ile	Pro	Ile	Asn	Val	Tyr		
			145			150				155					160		
AAG	CTG	CTG	GCA	GAG	GAC	TGG	CCA	TTT	GGA	GCT	GAG	ATG	TGT	AAG	CTG		765
Lys	Leu	Leu	Ala	Glu	Asp	Trp	Pro	Phe	Gly	Ala	Glu	Met	Cys	Lys	Leu		
				165					170				175				
GTG	CCT	TTC	ATA	CAG	AAA	GCC	TCC	GTG	GGA	ATC	ACT	GTG	CTG	AGT	CTA		813
Val	Pro	Phe	Ile	Gln	Lys	Ala	Ser	Val	Gly	Ile	Thr	Val	Leu	Ser	Leu		
			180					185					190				
TGT	GCT	CTG	AGT	ATT	GAC	AGA	TAT	CGA	GCT	GTT	GCT	TCT	TGG	AGT	AGA		861

Sub 1

Cys	Ala	Leu	Ser	Ile	Asp	Arg	Tyr	Arg	Ala	Val	Ala	Ser	Trp	Ser	Arg		
		195					200						205				
ATT	AAA	GGA	ATT	GGG	GTT	CCA	AAA	TGG	ACA	GCA	GTA	GAA	ATT	GTT	TTG		909
Ile	Lys	Gly	Ile	Gly	Val	Pro	Lys	Trp	Thr	Ala	Val	Glu	Ile	Val	Leu		
	210					215						220					
ATT	TGG	GTG	GTC	TCT	GTG	GTT	CTG	GCT	GTC	CCT	GAA	GCC	ATA	GGT	TTT		957
Ile	Trp	Val	Val	Ser	Val	Val	Leu	Ala	Val	Pro	Glu	Ala	Ile	Gly	Phe		
225					230					235					240		
GAT	ATA	ATT	ACG	ATG	GAC	TAC	AAA	GGA	AGT	TAT	CTG	CGA	ATC	TGC	TTG		1005
Asp	Ile	Ile	Thr	Met	Asp	Tyr	Lys	Gly	Ser	Tyr	Leu	Arg	Ile	Cys	Leu		
			245					250						255			
CTT	CAT	CCC	GTT	CAG	AAG	ACA	GCT	TTC	ATG	CAG	TTT	TAC	AAG	ACA	GCA		1053
Leu	His	Pro	Val	Gln	Lys	Thr	Ala	Phe	Met	Gln	Phe	Tyr	Lys	Thr	Ala		
			260					265						270			
AAA	GAT	TGG	TGG	CTG	TTC	AGT	TTC	TAT	TTC	TGC	TTG	CCA	TTG	GCC	ATC		1101
Lys	Asp	Trp	Trp	Leu	Phe	Ser	Phe	Tyr	Phe	Cys	Leu	Pro	Leu	Ala	Ile		
		275					280					285					
ACT	GCA	TTT	TTT	TAT	ACA	CTA	ATG	ACC	TGT	GAA	ATG	TTG	AGA	AAG	AAA		1149
Thr	Ala	Phe	Phe	Tyr	Thr	Leu	Met	Thr	Cys	Glu	Met	Leu	Arg	Lys	Lys		
		290				295					300						
AGT	GGC	ATG	CAG	ATT	GCT	TTA	AAT	GAT	CAC	CTA	AAG	CAG	AGA	CGG	GAA		1197

Subbi

Ser	Gly	Met	Gln	Ile	Ala	Leu	Asn	Asp	His	Leu	Lys	Gln	Arg	Arg	Glu	
305					310					315					320	
GTG	GCC	AAA	ACC	GTC	TTT	TGC	CTG	GTC	CTT	GTC	TTT	GCC	CTC	TGC	TGG	1245
Val	Ala	Lys	Thr	Val	Phe	Cys	Leu	Val	Leu	Val	Phe	Ala	Leu	Cys	Trp	
				325					330					335		
CTT	CCC	CTT	CAC	CTC	AGC	AGG	ATT	CTG	AAG	CTC	ACT	CTT	TAT	AAT	CAG	1293
Leu	Pro	Leu	His	Leu	Ser	Arg	Ile	Leu	Lys	Leu	Thr	Leu	Tyr	Asn	Gln	
				340				345						350		
AAT	GAT	CCC	AAT	AGA	TGT	GAA	CTT	TTG	AGC	TTT	CTG	TTG	GTA	TTG	GAC	1341
Asn	Asp	Pro	Asn	Arg	Cys	Glu	Leu	Leu	Ser	Phe	Leu	Leu	Val	Leu	Asp	
		355					360						365			
TAT	ATT	GGT	ATC	AAC	ATG	GCT	TCA	CTG	AAT	TCC	TGC	ATT	AAC	CCA	ATT	1389
Tyr	Ile	Gly	Ile	Asn	Met	Ala	Ser	Leu	Asn	Ser	Cys	Ile	Asn	Pro	Ile	
	370					375					380					
GCT	CTG	TAT	TTG	GTG	AGC	AAA	AGA	TTC	AAA	AAC	TGC	TTT	AAG	TCA	TGC	1437
Ala	Leu	Tyr	Leu	Val	Ser	Lys	Arg	Phe	Lys	Asn	Cys	Phe	Lys	Ser	Cys	
385					390					395					400	
TTA	TGC	TGC	TGG	TGC	CAG	TCA	TTT	GAA	GAA	AAA	CAG	TCC	TTG	GAG	GAA	1485
Leu	Cys	Cys	Trp	Cys	Gln	Ser	Phe	Glu	Glu	Lys	Gln	Ser	Leu	Glu	Glu	
				405				410					415			
AAG	CAG	TCG	TGC	TTA	AAG	TTC	AAA	GCT	AAT	GAT	CAC	GGA	TAT	GAC	AAC	1533

Subb<sup>1</sup>  
Lys Gln Ser Cys Leu Lys Phe Lys Ala Asn Asp His Gly Tyr Asp Asn  
420 425 430

TTC CGT TCC AGT AAT AAA TAC AGC TCA TCT TGAAAGAAGA ACTATTCACT 1583  
Phe Arg Ser Ser Asn Lys Tyr Ser Ser Ser  
435 440  
GTATTTTCATT TTCTTTATAT TGGACCGAAG TCATTAAAAC AAAATGAAAC ATTTGCCAAA 1643  
ACAAAACAAA AAACATATGTA TTTGCACAGC ACACTATTAA AATATTAAGT GTAATTATTT 1703  
TAACACTCAC AGCTACATAT GACATTTTAT GAGCTGTTTA CGGCATGGAA AGAAAATCAG 1763  
TGGGAATTAA GAAAGCCTCG TCGTGAAAGC ACTTAATTTT TTACAGTTAG CACTTCAACA 1823  
TAGCTCTTAA CAACTTCCAG GATATTCACA CAACACTTAG GCTTAAAAT GAGCTCACTC 1883  
AGAATTTCTA TTCTTTCTAA AAAGAGATTT ATTTTAAAT CAATGGGACT CTGATATAAA 1943  
GGAAGAATAA GTCACTGTAA AACAGAACTT TTAAATGAAG CTAAATTAC TCAATTTAAA 2003  
ATTTTAAAAT CCTTTAAAAC AACTTTTCAA TTAATATTAT CACACTATTA TCAGATTGTA 2063  
ATTAGATGCA AATGAGAGAG CAGTTTAGTT GTTGCATTTT TCGGACACTG GAAACATTTA 2123  
AATGATCAGG AGGGAGTAAC AGAAAGAGCA AGGCTGTTTT TGAAAATCAT TACACTTTCA 2183  
CTAGAAGCCC AAACCTCAGC ATTCTGCAAT ATGTAACCAA CATGTCACAA AGAAGCAGCA 2243  
TGTAACAGAC TGGCACATGT GCCAGCTGAA TTAAAATAT AATACTTTTA AAAAGAAAAT 2303  
TATTACATCC TTTACATTCA GTTAAGATCA AACCTCACAA AGAGAAATAG AATGTTTGAA 2363  
AGGCTATCCC AAAAGACTTT TTTGAATCTG TCATTCACAT ACCCTGTGAA GACAATACTA 2423

Subb 17

009575760

TCTACAATTT	TTTCAGGATT	ATTAAAATCT	TCTTTTTTCA	CTATCGTAGC	TTAAACTCTG	2483
TTTGGTTTTG	TCATCTGTAA	ATACTTACCT	ACATACACTG	CATGTAGATG	ATTAAATGAG	2543
GGCAGGCCCT	GTGCTCATAG	CTTTACGATG	GAGAGATGCC	AGTGACCTCA	TAATAAAGAC	2603
TGTGAACTGC	CTGGTGCAGT	GTCACATGA	CAAAGGGGCA	GGTAGCACCC	TCTCTCACCC	2663
ATGCTGTGGT	TAAAATGGTT	TCTAGCATAT	GTATAATGCT	ATAGTTAAAA	TACTATTTTT	2723
CAAAATCATA	CAGATTAGTA	CATTTAACAG	CTACCTGTAA	AGCTTATTAC	TAATTTTTGT	2783
ATTATTTTTG	TAAATAGCCA	ATAGAAAAGT	TTGCTTGACA	TGGTGCTTTT	CTTTCATCTA	2843
GAGGCAAAAC	TGCTTTTTGA	GACCGTAAGA	ACCTCTTAGC	TTTGTGCGTT	CCTGCCTAAT	2903
TTTTATATCT	TCTAAGCAAA	GTGCCTTAGG	ATAGCTTGGG	ATGAGATGTG	TGTGAAAGTA	2963
TGTACAAGAG	AAAACGGAAG	AGAGAGGAAA	TGAGGTGGGG	TTGGAGGAAA	CCCATGGGGA	3023
CAGATTCCCA	TTCTTAGCCT	AACGTTTCGT	ATTGCTTCGT	CACATCAATG	CAAAGGTCC	3083
TGATTTTGTT	CCAGCAAAAC	ACAGTGCAAT	GTTCTCAGAG	TGACTTTCGA	AATAAATTGG	3143
GCCCAAGAGC	TTTAACTCGG	TCTTAAAATA	TGCCCAAATT	TTTACTTTGT	TTTTCTTTTA	3203
ATAGGCTGGG	CCACATGTTG	GAAATAAGCT	AGTAATGTTG	TTTTCTGTCA	ATATTGAATG	3263
TGATGGTACA	GTAAACCAAA	ACCCAACAAT	GTGGCCAGAA	AGAAAGAGCA	ATAATAATTA	3323
ATTCACACAC	CATATGGATT	CTATTTATAA	ATCACCCACA	AACTTGTTCT	TTAATTTTCT	3383
CCCAATCACT	TTTTCAGAGG	CCTGTTATCA	TAGAAGTCAT	TTTAGACTCT	CAATTTTAAA	3443
TTAATTTTGA	ATCACTAATA	TTTTCACAGT	TTATTAATAT	ATTTAATTTT	TATTTAAATT	3503

Sub B1

TTAGATTATT TTTATTACCA TGTACTGAAT TTTTACATCC TGATACCCTT TCCTTCTCCA 3563

TGTCAGTATC ATGTTCTCTA ATTATCTTGC CAAATTTTGA AACTACACAC AAAAAGCATA 3623

CTTGCAATTAT TTATAATAAA ATTGCAITCA GTGGCTTTTT AAAAAAATG TTTGATTCAA 3683

AACTTTAACA TACTGATAAG TAAGAAACAA TTATAATTTC TTTACATACT CAAAACCAAG 3743

ATAGAAAAAG GTGCTATCGT TCAACTTCAA AACATGTTTC CTAGTATTAA GGACTTTAAT 3803

ATAGCAACAG ACAAATTAT TGTTAACATG GATGTTACAG CTCAAAAGAT TTATAAAAGA 3863

TTTTAACCTA TTTTCTCCCT TATTATCCAC TGCTAATGTG GATGTATGTT CAAACACCTT 3923

TTAGTATTGA TAGCTTACAT ATGGCCAAAG GAATACASTT TATAGCAAAA CATGGGTATG 3983

CTGTAGCTAA CTTTATAAAA GTGTAATATA ACAATGTAAA AAATTATATA TCTGGGAGGA 4043

TTTTTTGGTT GCCTAAAGTG GCTATAGTTA CTGATTTTTT ATTATGTAAG CAAAACCAAT 4103

AAAAATTTAA GTTTTTTTAA CAACTACCTT ATTTTTCCT GTACAGACAC TAATTCATTA 4163

AATACTAATT GATTGTTTAA AAGAAATATA AATGTGACAA GTGGACATTA TTTATGTTAA 4223

ATATACAATT ATCAAGCAAG TATGAAGTTA TTCAATTAAA ATGCCACATT TCTGGTCTCT 4283

GGGAAAAAAA AAAAAAAA 4301